

# Reliability and Validity of the Aggregate Method of Determining Number of Cigarettes Smoked Per Day

Peter W. Gariti, Ph.D., Arthur I. Alterman, Ph.D.  
Ronald N. Ehrman, Ph.D., Helen M. Pettinati, Ph.D.

*The authors evaluated the reliability of two pretreatment assessments (screening and intake) of cigarettes smoked per day (CPD) by the commonly used aggregate method. The validity of the aggregate method was also determined by comparison with results of the timeline followback (TLFB) method for the identical periods. The study participants were 49 outpatients undergoing nicotine patch treatment. The reliability of the two aggregate method evaluations of CPD was quite high by Pearson product-moment correlation ( $r$ ) and good when based on the intraclass correlation. Correspondence between the CPD assessments based on the aggregate and TLFB methods for the two time-points ranged from fair (screening) to good (intake). Overall, the study findings indicate that the aggregate method provides reasonably consistent data. (Am J Addict 1998; 7:283-287)*

Accurate measurement of smoking is necessary for evaluating the effectiveness of cessation treatment. Obtaining an accurate and reliable estimate of pretreatment (as well as during- and posttreatment) smoking patterns is an important factor not only in measuring outcome over time but also in accurately classifying subjects into light, medium, and heavy smokers, determining recommended patch dosage, monitoring patch adherence, and comparing nicotine/cotinine replacement values with baseline level of smoking.

Given that biological markers of choice (plasma, urine, or saliva cotinine) used in most clinical trials are neither economical nor adequate reflections of the amount of smoking over extended periods of time, assessment of the amount of smoking is largely dependent on the respondent's self-report of cigarette use. An obvious question that can be raised is how we know that what is reported during screenings or just before the start of treatment is an accurate reflection of actual smoking behavior? Arguably, a potential patient may deliberately over- or

Received December 17, 1997; accepted April 3, 1998. From The Philadelphia Veterans Affairs Medical Center/University of Pennsylvania School of Medicine, Philadelphia, PA. Address correspondence to Dr. Gariti, Treatment Research Center, 3900 Chestnut Street, Philadelphia, PA 19104.

underestimate the number of cigarettes smoked because of perceived demand characteristics related to treatment entry. Alternatively, a respondent may simply provide a gross estimate rounded off to the nearest pack or half-pack.<sup>1</sup> Hence, the reported number of cigarettes smoked initially may not necessarily reflect the actual amount being smoked at the beginning of treatment.

The method for determining baseline cigarette usage in most smoking-cessation patch studies is the "aggregate" method. Using this method, potential recruits are typically asked the number of cigarettes currently being smoked, which elicits a summary estimate of the number of cigarettes smoked per day (CPD). The apparent assumption made in most studies is that there is little variability in the level of smoking before the start of actual treatment. It frequently has been reported that except for attempts at quitting or other intervening circumstances, such as a hospitalization, there is minimal fluctuation in smoking level after the first 5 years of smoking initiation.<sup>2</sup> However, the observation that smokers tend to reduce their habit during the work week in response to an increase in the number of work sites that are smoke-free contradicts this assumption. Therefore, a "quantity-frequency" (QF) method that evaluates the number of days smoked as well as the number of cigarettes smoked per given day over an extended time-frame (i.e., the past month or longer) may provide a more accurate profile of CPD, particularly in persons with a variable pattern of use.

Sobell and Sobell<sup>3</sup> have developed a highly reliable variation of the quantity-frequency method for obtaining patient self-report of alcohol use retrospectively over extended time periods. Timeline followback (TLFB) reports have consistently demonstrated high test-retest reliability across multiple population types (normal drinkers and alcoholic subjects) and have been reasonably congruent with collateral supportive data<sup>4</sup> and single-sample biological testing,

namely liver-function testing (SGOT/SGPT).<sup>3</sup> In a recent application of TLFB to the assessment of substance abuse, Ehrman and Robbins<sup>5</sup> reported that TLFB estimates of cocaine and heroin use over a 6-month period correlated highly with program-based weekly qualitative urine specimens in a sample of methadone-maintenance patients.

Given the foregoing considerations, two basic questions were evaluated in this study: 1) the extent to which two pretreatment aggregate assessments of CPD collected 1 month apart correlate with each other (reliability); and 2) the extent to which CPD data obtained using the two aggregate measures correspond with a subsequent TLFB assessment of CPD for the same two pretreatment time-points (validity).

## METHODS

### Research Participants

Participants consisted of 49 patients (30 female, 19 male) recruited and randomized into a smoking-cessation study at a university-based substance-abuse outpatient setting. The participants, chosen by telephone screening, were physically healthy and mentally stable men and nonpregnant women between the ages of 18 and 65, who reported smoking at least one pack of cigarettes daily for at least the past month, met DSM-IV criteria for nicotine dependence on the basis of a semistructured diagnostic interview, and reported at least one previous failed attempt at smoking cessation. Prospective participants were excluded if they had any medical condition that would preclude the use of the patch (e.g., unstable cardiovascular disease, allergies to the patch, peptic ulcer), had serious cognitive disorders, were currently psychotic or expressed current homicidal or suicidal ideation, met DSM-IV criteria for current non-nicotine substance abuse/dependence within the past 6 months, or were currently using cocaine or nonprescribed amphetamine.

Participants were recruited from a number of sources, including university campus notices, local newspaper advertisements, and word-of-mouth. Written informed consent was obtained after subjects received a complete study description and passed a consent form quiz.

The average subject was  $43 \pm 9$  years of age; 61% were women, and 65% were white (33% African-American). On average, participants had completed  $15 \pm 3$  years of school; 33% were currently married; and 80% currently employed. They had smoked for an average of  $23 \pm 9$  years and had made  $7 \pm 11$  previous attempts to quit smoking.

#### Procedures

Because recruitment notices did not specify study requirements, research candidates were unaware of inclusion/exclusion criteria regarding smoking behavior before they were screened. No attempt was made to have subjects alter cigarette use before the actual start of patch treatment, nor were they told that their responses would be cross-checked for consistency.

Research technicians questioned subjects about their cigarette use on three separate occasions: 1) during a technician-administered structured telephone screening that used the aggregate method; 2) at intake approximately 1 month later by a technician as well as a psychiatrist separately confirming responses to a standardized smoking-history questionnaire using the aggregate method; and 3) at a technician-administered TLFB interview that took place at initiation of patch treatment approximately 1 month after intake. Fifteen of the 49 subjects were evaluated by different technicians at the three time-points, whereas 34 of 49 were evaluated each time by the same technician.

About 25% of the subjects at the telephone screening (but not at intake) specified a range of CPD when the aggregate method was used. In these cases, the mid-point value was selected a priori. The TLFB

procedure used an adapted script based on Sobell and Sobell's instructions for completing a timeline drinking calendar.<sup>5</sup> Participants were asked to provide their best estimate of daily cigarette usage across the preceding 6-month period. Anchoring-points, such as weekdays, weekends, holidays, paydays, days off, and the recall of major life events or personal events were elicited to aid the recall of past smoking for the TLFB. TLFB responses were then averaged, using the dates of the initial telephone screen, intake, and start of treatment as the anchoring-points for determining the average number of cigarettes smoked 1 month before each of the aforementioned anchoring-points.

#### RESULTS

The degree of association between the two estimates of CPD using the aggregate and TLFB methods for each of these time periods was evaluated using both the Pearson product-moment correlation ( $r$ ), and the intraclass correlation (ICC) formula developed by Lin.<sup>6</sup> Although  $r$  provides information on the degree of ordinal relationship, it provides no information on absolute agreement. The ICC takes absolute values into account and provides a measure of extent of exact agreement. The ICC is clearly a more accurate indicator of degree of relationship.<sup>7</sup>

#### Reliability of the Aggregate Method

The correlation between the CPD at screening was compared with that obtained at intake. The Pearson  $r$  was 0.92 ( $P < 0.0005$ ), and the ICC was 0.67. According to Cicchetti,<sup>7</sup> an ICC between 0.60 and 0.74 is indicative of good agreement. The mean CPD at these two time-points was  $28.42 \pm 10.3$  and  $27.59 \pm 10.2$ , respectively (paired  $t_{48} = 1.43$ ;  $P = 0.12$ ).

## Validity of the Aggregate Method: Comparison With TLFB

**Screening.** The relationship between the aggregate and TLFB assessments of CPD at screening was first compared. The Pearson  $r$  was 0.81 ( $P < 0.0005$ ), and ICC was 0.42. This latter value should be interpreted as indication of, at most, a fair degree of correspondence between the two measurement approaches.<sup>6</sup> The mean for the aggregate method was  $28.42 \pm 10.3$ , and mean for the TLFB method was  $27.01 \pm 10.2$ ; these did not differ significantly (paired  $t_{48} = 1.57$ ;  $P = 0.12$ ).

**Intake.** The Pearson  $r$  comparing aggregate and TLFB CPDs was 0.85 ( $P < 0.0005$ ), and the ICC was 0.60. The value of 0.60 for the ICC suggests good correspondence between the two methods of assessment. The means for the two methods were  $27.59 \pm 10.2$  and  $26.86 \pm 9.8$ , respectively. The paired-sample  $t$ -test showed no significant differences ( $t_{48} = 0.94$ ;  $P = 0.35$ ).

The findings described above were essentially the same for those cases in which the same interviewer performed the various assessments and those in which different interviewers performed these assessments.

## DISCUSSION

The study findings indicate a reasonably high level of within-subject consistency between assessments at different time-points in reporting CPD by use of the aggregate method. Comparison of the findings for the aggregate method with the more precise and presumably more accurate TLFB method revealed a moderate or high degree of association, depending on the measure of association used.

The relationship between the aggregate and TLFB methods was determined to be high when absolute level was not taken into account (Pearson product-moment correlation [ $r$ ]), but was only fair-to-good when the

absolute level was considered (intraclass correlation [ICC]), that is, when exact agreement was required. Thus, the evidence suggests a good degree of validity for the aggregate method, but also points to some limits in its validity, given that correspondence between the two assessment approaches was not excellent. These data reinforce the conclusion that although the aggregate method may provide a satisfactory rough approximation of amount of smoking, especially at the level of group analysis, its results are not as satisfactory for characterizing absolute changes in individual smoking from time to time.

More detailed examination of the TLFB data revealed that smoking was significantly elevated on Saturdays, as compared with weekdays. At the same time, smoking was not especially elevated on Sundays. The variation between Saturday and weekday smoking should be further evaluated because it suggests that patients may be more vulnerable to relapse at that time. This is information that cannot be extracted from the more global aggregate method.

Overall, the study findings indicate that the aggregate method provides reasonably consistent estimates of CPD. It may be the method of choice when time and cost considerations are important because of its relative brevity, given that the TLFB takes 5 to 20 minutes to administer. By contrast, TLFB may offer more detailed and presumably more accurate data on CPD time course and may be preferable when its greater cost and inconvenience are not critical considerations.

An important albeit secondary concern of this research was whether there was indication of decreases in patients' smoking in the week immediately before treatment. Our analyses of the relationship between TLFB at screening and 1 week before the start of treatment indicated little variation ( $< 1$  CPD) in self-reported CPD as subjects approached their "QUIT DATE."

It should be emphasized that both

methods, the aggregate and TLFB, rely on patients' self-report. Although the generally satisfactory degree of correspondence between the two methods suggests that patients report CPD relatively accurately, the ultimate validity of these self-report data awaits the development of more adequate objective assessment methods.

Another important possible limitation

of the findings is that the data were based on individuals who smoked at least 20 cigarettes daily. The same results may not be obtained for more moderate smokers.

*This study was supported by Grant #DA 10070 from the National Institute on Drug Abuse and by an NIDA Center Grant. The authors thank Gary Luck for assistance in analysis of the data.*

#### References

1. Kleges RC, Debon M, Ray JW: Are self-reports of smoking rate biased? evidence from the second national health and nutrition examination survey. *J Clin Epidemiol* 1995; 10:1225-1233
2. U.S. Department of Health and Human Services: The Health Consequences of Smoking: Nicotine Addiction: A Report of the Surgeon General, Washington, DC, U.S. Government Printing Office, 1988
3. Sobell LC, Sobell MB: Timeline follow-back: a technique for assessing self-reported alcohol consumption, in *Measuring Alcohol Consumption: Psychosocial and Biological Methods*. Edited by Litten RZ, Allen J. New Jersey, Humana Press, 1991, pp 2-40
4. Sobell LC, Sobell MB, Leo GL, et al: Reliability of a timeline method: assessing normal drinkers' report of recent drinking and a comparative evaluation across several populations. *Br J Addict* 1988; 83:393-402
5. Ehrman RN, Robbins SJ: Reliability and validity of six-month timeline reports of cocaine and heroin use in a methadone population. *J Consult Clin Psychol* 1994; 62:1-8
6. H-Kuei Lin L: A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; 45:255-268
7. Cicchetti D: Guidelines, criteria, and rules for evaluating normed and standardized assessment instruments. *Psychol Assess* 1994; 6:284-290